New integer and Bilevel Formulations for the k-Vertex Cut Problem

Ivana Ljubić[•],

joint work with F. Furini[◦], E. Malaguti[*] and P. Paronuzzi[*]

[*]DEI "Guglielmo Marconi", University of Bologna

[◦]IASI-CNR, Rome

[•]ESSEC Business School, Paris

SPOC22 Meeting, Oct 30 2020, online edition

Outline

1. Problem setting and motivation
2. Compact Model
3. Representative Formulation
4. Bilevel approach
5. A Hybrid Approach?
6. Computational experiments

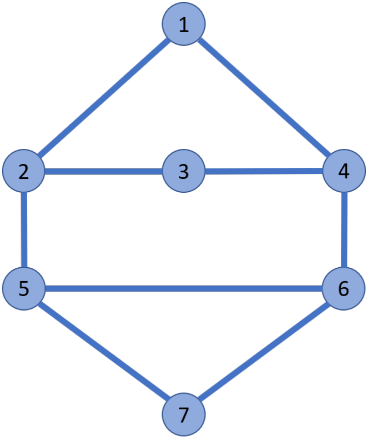
Based on the paper

Fabio Furini, Ivana Ljubić, Enrico Malaguti, Paolo Paronuzzi:
On integer and bilevel formulations for the k-vertex cut problem.
Math. Program. Comput. 12(2): 133-164 (2020)

Problem setting and motivation

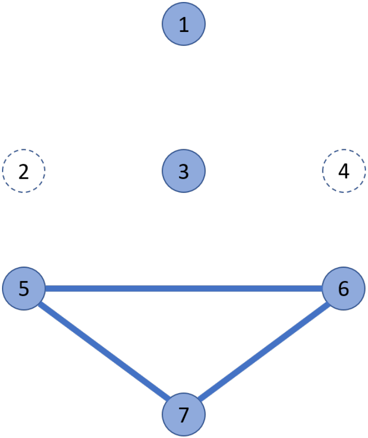
Problem setting

A **k-vertex cut** is a subset of vertices whose removal disconnects the graph in at least k (not-empty) components.

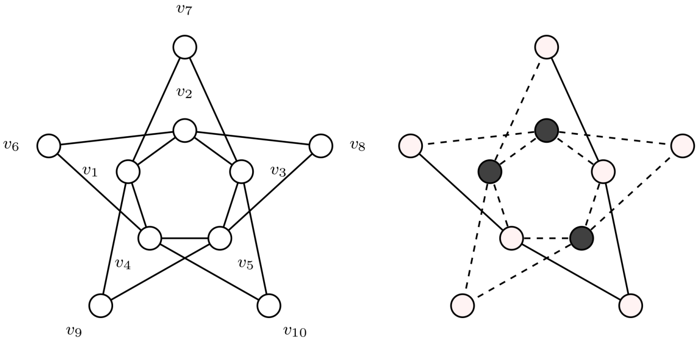


Problem setting

Example of a 3-vertex cut:



Another example of a 3-vertex cut:



The k-Vertex Cut Problem

Definition

Given an undirected graph $G = (V, E)$ with vertex weights w_v, $v \in V$, and a integer $k \geq 2$, find a subset of vertices of **minimum weight** whose removal disconnects G in at least k (not-empty) components.

Motivation

- **Family of Critical Node Detection Problems** (M. Lalou, M. A. Tahraoui, and H. Kheddouci. The critical node detection problem in networks: A survey. Computer Science Review, 2018);
  - **Defender-Attacker** model: k-vertex cut are vertices to be **defended** (protected, vaccinated)
  - **Attacker-Defender** model: k-vertex cut are vertices to be **attacked**
- **Decomposition method for linear equation systems**: vertices are columns of a matrix, and an edge connects two vertices if there are two non-zero entries in the same raw.

Problem Complexity

Vertex k-cut

- $k = 2$: polynomial (Ben Ameur & Biha, 2012).
- $k \geq 3$: NP-hard (Berger et al, 2014), even for a fixed value of k (Cornaz et al, 2019)

Edge k-cut

- For any fixed value of k: polynomial (Goldschmidt & Hochbaum, 1994).
- 2-approximation algorithms exist.

Compact Model

Compact formulation

We associate a binary variable y_v^i to all vertices $v \in V$ and for all integers $i \in K$, such that:

$$y_v^i = \begin{cases} 1 & \text{if vertex } v \text{ belongs to component } i \\ 0 & \text{otherwise} \end{cases} \qquad i \in K, v \in V.$$

If $\sum_{i \in K} y_v^i = 0 \quad \Rightarrow$ v belongs to the k-vertex cut.

Compact Model

Compact ILP formulation for k-Vertex-Cut Problem:

$$\min \sum_{v \in V} w_v - \sum_{i \in K} \sum_{v \in V} w_v y_v^i$$

$$\sum_{i \in K} y_v^i \leq 1 \qquad v \in V,$$

$$y_u^i + \sum_{j \in K \setminus \{i\}} y_v^j \leq 1 \qquad i \in K, uv \in E,$$

$$\sum_{v \in V} y_v^i \geq 1 \qquad i \in K,$$

$$y_v^i \in \{0, 1\} \qquad i \in K, v \in V.$$

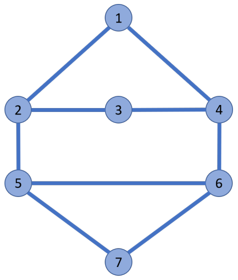
Drawbacks: LP-optimal solution is zero (set all $y_v^i = 1/k$), symmetries, etc.

Representative Formulation

Representative formulation

Observation

A graph $G = (V, E)$ admits a k-vertex cut if and only if its *stability number* $\alpha(G)$ is at least k.



Representative formulation

In the **Multi-Terminal Vertex Separator** problem a set of representative vertices for each component are given.
Similar model is proposed in Y. Magnouche's Ph.D. thesis (2017).

We introduce a set of binary variable to select which vertices are representative:

$$z_v = \begin{cases} 1 & \text{if vertex } v \text{ is the representative of a component} \\ 0 & \text{otherwise} \end{cases} \qquad v \in V$$

and we use the same set of binary variables denoting whether a vertex is in the k-vertex cut:

$$x_v = \begin{cases} 1 & \text{if vertex } v \text{ is in the } k\text{-vertex cut} \\ 0 & \text{otherwise} \end{cases} \qquad v \in V$$

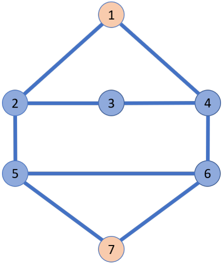
Representative formulation

The *Representative Formulation* reads as follows:

$$
\begin{aligned}
\min \sum_{v \in V} x_v & & \\
\sum_{v \in V} z_v = k & & v \in V, \\
z_u + z_v \leq 1 & & uv \in E, \\
\sum_{w \in V(P) \setminus \{u,v\}} x_w \geq z_u + z_v - 1 & & u, v \in V, P \in \Pi_{uv}, uv \notin E \\
x_v, z_v \in \{0, 1\} & & v \in V.
\end{aligned}
$$

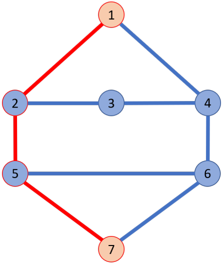
Representative formulation: Path Inequalities

$$\sum_{w \in V(P) \setminus \{u,v\}} x_w \geq z_u + z_v - 1$$



Representative formulation: Path Inequalities

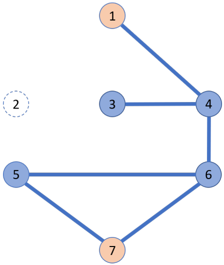
$$\sum_{w \in V(P) \setminus \{u,v\}} x_w \geq z_u + z_v - 1$$



$$x_2 + x_5 \geq z_1 + z_7 - 1$$

Representative formulation: Path Inequalities

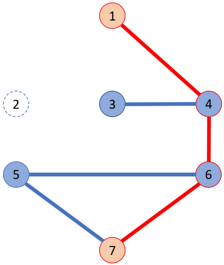
$$\sum_{w \in V(P) \setminus \{u,v\}} x_w \geq z_u + z_v - 1$$



$$x_2 + x_5 \geq z_1 + z_7 - 1$$

Representative formulation: Path Inequalities

$$\sum_{w \in V(P) \setminus \{u,v\}} x_w \geq z_u + z_v - 1$$

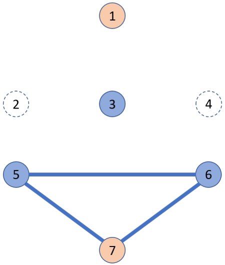


$$x_2 + x_5 \geq z_1 + z_7 - 1$$

$$x_4 + x_6 \geq z_1 + z_7 - 1$$

Representative formulation: Path Inequalities

$$\sum_{w \in V(P) \setminus \{u,v\}} x_w \geq z_u + z_v - 1$$



$$x_2 + x_5 \geq z_1 + z_7 - 1$$

$$x_4 + x_6 \geq z_1 + z_7 - 1$$

Separation of Path Inequalities

Given a solution $x^*, z^* \in [0,1]^V$, the separation problem asks for finding a pair of vertices u, v such that there is a path $P^* \in \Pi_{uv}$ with

$$z_u + z_v > \sum_{w \in V(P^*) \setminus \{u,v\}} x_w - 1.$$

We can search for such a path in polynomial time by solving a shortest path problem (for each pair of not adjacent vertices) on graph G, where we define the length of each edge $(u, v) \in E$ as:

$$l_{uv} = \frac{x_u^* + x_v^*}{2}$$

Representative formulation

- Valid constraints in polynomial number:

$$\begin{aligned} x_u + z_u &\leq 1 & u \in V, \\ z_u + \sum_{v \in N(u)} z_v &\leq 1 + (\deg(u) - 1) x_u & u \in V. \end{aligned}$$

- Strengthened Path Inequalities:

$$\sum_{w \in V(P) \setminus \{u,v\}} x_w \geq z_u + z_v + \sum_{w \in V(P) \setminus \{u,v\}} z_w - 1$$

- Clique-Path Inequalities... Each z on the RHS is replaced by a clique...

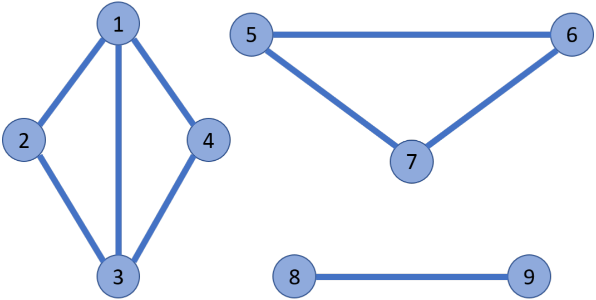
Bilevel approach

Bilevel approach

Property

A graph G has at least k (not empty) components if and only if any cycle-free subgraph of G contains at most $|V| - k$ edges.

Example with $|V| = 9$ and $k = 3$:

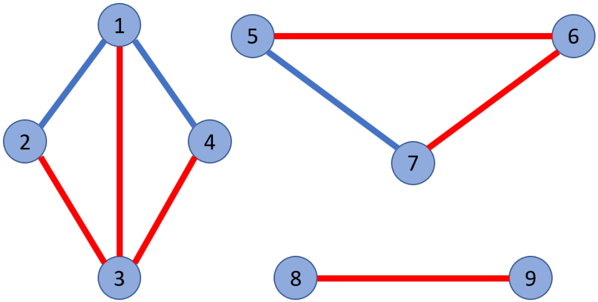


Bilevel approach

Property

A graph G has at least k (not empty) components if and only if any cycle-free subgraph of G contains at most $|V| - k$ edges.

Example with $|V| = 9$ and $k = 3$:



Bilevel approach

The k-vertex cut problem can be seen as a Stackelberg game:

- the leader searches the smallest subset of vertices V_0 to delete;
- the follower **maximizes the size of the cycle-free subgraph** on the residual graph.

Property

The solution $V_0 \subset V$ of the leader is feasible if and only if the value of the **optimal follower's response** *(i.e., the size of the maximum cycle-free subgraph in the remaining graph)* **is at most $|V| - |V_0| - k$.**

Bilevel approach

The leader decisions:

$$x_v = \begin{cases} 1 & \text{if vertex } v \text{ is in the } k\text{-vertex cut} \\ 0 & \text{otherwise} \end{cases} \qquad v \in V$$

For the decisions of the follower, we use additional binary variables associated with the edges of G:

$$e_{uv} = \begin{cases} 1 & \text{if edge } uv \text{ is selected to be in the cycle-free subgraph} \\ 0 & \text{otherwise} \end{cases} \qquad uv \in E$$

Bilevel approach

The Bilevel ILP formulation of the k-vertex cut problem reads as follows:

$$\min \sum_{v \in V} x_v$$
$$\Phi(x) \leq |V| - \sum_{v \in V} x_v - k$$
$$x_v \in \{0,1\} \qquad v \in V.$$

- $\Phi(x)$ is the optimal solution value of the follower subproblem for a given x.
- **Value Function Reformulation**.
- Value function $\Phi(x)$ is neither convex, nor concave, nor connected...

How do we calculate $\Phi(x)$?

For a solution x^* of the leader, which denotes a set V_0 of interdicted vertices, the follower's subproblem is:

$$\begin{aligned}
\Phi(x^*) = \quad & \max \sum_{uv \in E} e_{uv} & \\
& e(S) \leq |S| - 1 & S \subseteq V, S \neq \emptyset, \\
& e_{uv} \leq 1 - x^*_u & uv \in E, \\
& e_{uv} \leq 1 - x^*_v & uv \in E, \\
& e_{uv} \in \{0, 1\} & uv \in E.
\end{aligned}$$

Bilevel approach

We can prove that the follower's subproblem is equivalently restated as:

$$\Phi(x^*) = \quad \max \sum_{uv \in E} z_{uv}(1 - x^*_u - x^*_v)$$

$$z(S) \leq |S| - 1 \qquad S \subseteq V, S \neq \emptyset$$

$$z_{uv} \in \{0, 1\} \qquad uv \in E.$$

- **Convexification** of the value function $\Phi(x)$

Bilevel approach

Since the space of feasible solutions of the redefined follower subproblem does not depend on the leader anymore, the non-linear constraint from the BILP formulation:

$$\Phi(x) \leq |V| - \sum_{v \in V} x_v - k$$

can now be replaced by the following exponential family of inequalities:

$$\sum_{uv \in E(T)} (1 - x_u - x_v) \leq |V| - \sum_{v \in V} x_v - k \qquad T \in \mathcal{T}$$

where $\mathcal{T}$ denote the **set of all cycle-free subgraphs** of G.

Natural Formulation

The following single-level formulation, denoted as *Natural Formulation*, is a valid model for the k-vertex cut problem:

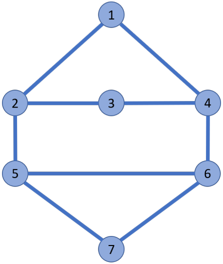
$$\min \sum_{v \in V} w_v x_v$$

$$\sum_{v \in V} [\deg_T(v) - 1] x_v \geq k - |V| + |E(T)| \qquad T \in \mathcal{T},$$

$$x_v \in \{0, 1\} \qquad v \in V.$$

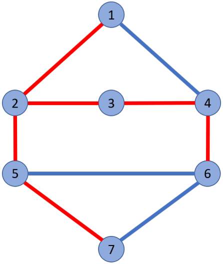
Natural formulation

$$\sum_{v \in V} [\deg_T(v) - 1] x_v \geq k - |V| + |E(T)|$$



Natural formulation

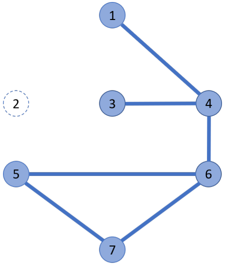
$$\sum_{v \in V} [\deg_T(v) - 1] x_v \geq k - |V| + |E(T)|$$



$$2x_2 + x_4 + x_5 \geq 2$$

Natural formulation

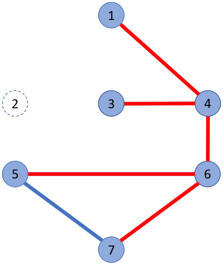
$$\sum_{v \in V} [\deg_T(v) - 1] x_v \geq k - |V| + |E(T)|$$



$$2x_2 + x_4 + x_5 \geq 2$$

Natural formulation

$$\sum_{v \in V} [\deg_T(v) - 1] x_v \geq k - |V| + |E(T)|$$

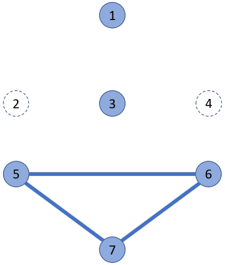


$$2x_2 + x_4 + x_5 \geq 2$$

$$-x_2 + 2x_4 + 2x_6 \geq 1$$

Natural formulation

$$\sum_{v \in V} [\deg_T(v) - 1] x_v \geq k - |V| + |E(T)|$$



$$2x_2 + x_4 + x_5 \geq 2$$

$$-x_2 + 2x_4 + 2x_6 \geq 1$$

Separation procedure

Let x^* be the current solution. We define edge-weights as

$$w^*_{uv} = 1 - x^*_u - x^*_v, \quad uv \in E$$

and search for the maximum-weighted cycle-free subgraph in G. Let W^* denote the weight of the obtained subgraph; if $W^* > |V| - k - \sum_{v \in V} x^*_v$, we have detected a violated inequality.

The separation procedure can be performed in polynomial time:

- adaptation of Kruskal's algorithm for minimum-spanning trees (fractional points), or
- BFS (integer points) on the graph from which $x_v = 1$ vertices are removed. Extended to spanning subgraphs (dominating cuts).

A Hybrid Approach?

Model Strength

Theorem

For $k \leq n/2$ we have:

$$v_{LP}(NAT) \leq v_{LP}(REP)$$

and there exist instances where the strict inequality holds.

Natural Formulation + Representative Constraints

$$\sum_{v \in V} z_v = k$$

$$z_u + z_v \leq 1 \qquad uv \in E,$$

$$x_u + z_u \leq 1 \qquad u \in V,$$

$$z_u + \sum_{v \in N(u)} z_v \leq 1 + (\deg(u) - 1)x_u \qquad u \in V.$$

Computational experiments

Computational Experiments

We considered two sets of graph instances from the 2nd DIMACS and 10th DIMACS challenges.
For all the instances we tested four different values of k (5, 10, 15, 20).

Compared Methods (time limit of 1 hour):

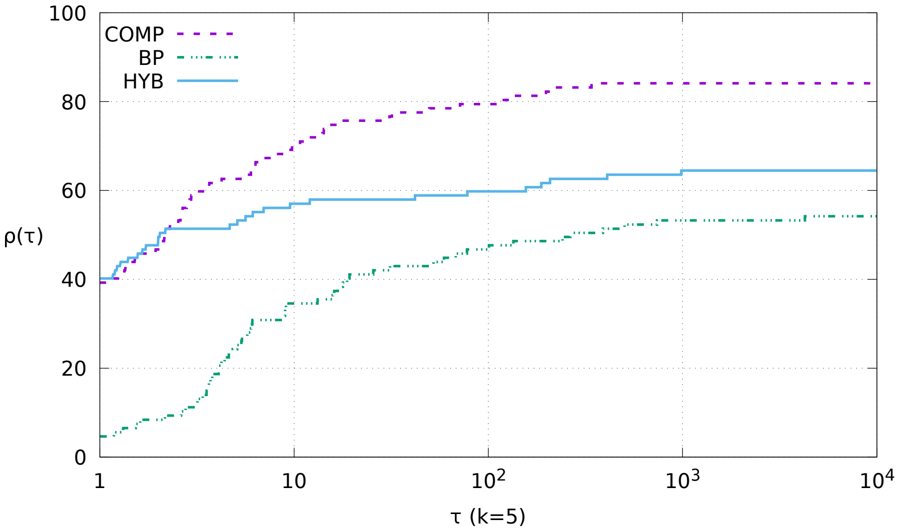
- **COMP:** *Compact* model (solved by CPLEX 12.7.1);
- **BP:** State-of-the-art Branch-and-Price solving an *Extended* formulation (*Cornaz, D., Furini, F., Lacroix, M., Malaguti, E., Mahjoub, A. R., & Martin, S. The Vertex k-cut Problem, Discrete Optimization, 2018.*);
- **HYB:** Hybrid approach

Comparing REP, NAT and HYB Formulation

k		REP	REP_{lp}	NAT	NAT_s	HYB
	Opt. (out of 51)	29	27	33	34	35
	Avg Time	243.87	161.32	101.35	130.76	193.63
5	Avg Nodes	77651	35418	74	77	45
	LP Avg Gap	94.19	74.27	42.39	41.98	41.96
	LP Avg Time	0.02	1.91	8.24	7.89	17.32
	Opt. (out of 41)	20	23	29	30	32
	Avg Time	191.59	413.46	87.08	138.92	177.33
10	Avg Nodes	39599	56357	21	23	23
	LP Avg Gap	86.31	56.20	31.82	31.80	31.84
	LP Avg Time	0.05	5.99	18.60	10.60	14.10
	Opt. (out of 38)	22	24	33	32	33
	Avg Time	138.69	162.92	327.88	208.37	87.51
15	Avg Nodes	57776	21875	57	38	19
	LP Avg Gap	77.82	47.42	23.96	23.97	23.98
	LP Avg Time	0.12	113.39	12.93	6.92	7.24
	Opt. (out of 36)	18	22	31	32	32
	Avg Time	133.50	432.51	147.68	185.29	145.05
20	Avg Nodes	45084	43520	53	62	25
	LP Avg Gap	69.89	41.85	21.25	21.24	21.25
	LP Avg Time	0.27	43.83	13.15	6.12	7.21
	Total Opt. (out of 166)	89	96	126	128	132
	Total Avg Time	183.80	288.59	168.80	165.71	151.37
	Total Avg Nodes	57600	38905	52	51	28
	Total Avg LP Gap	83.23	56.63	30.98	30.84	30.85

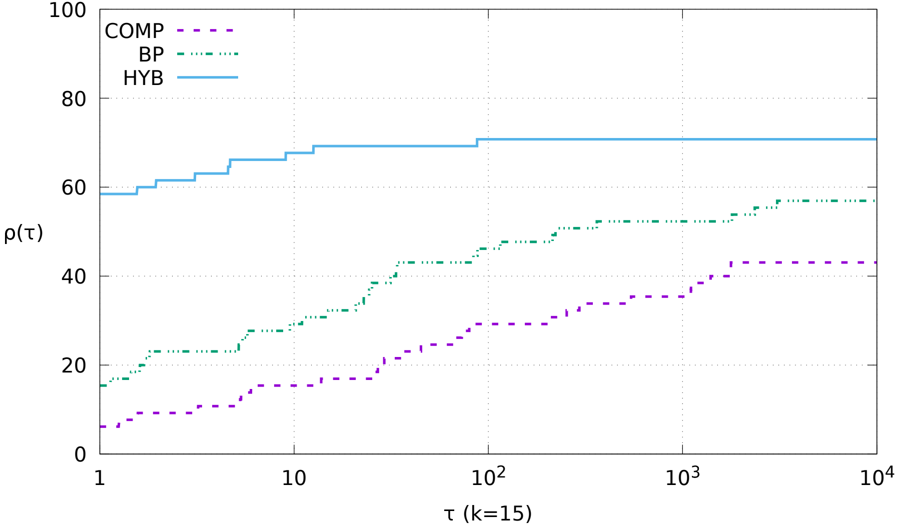
Computational Experiments

Case with $k = 5$



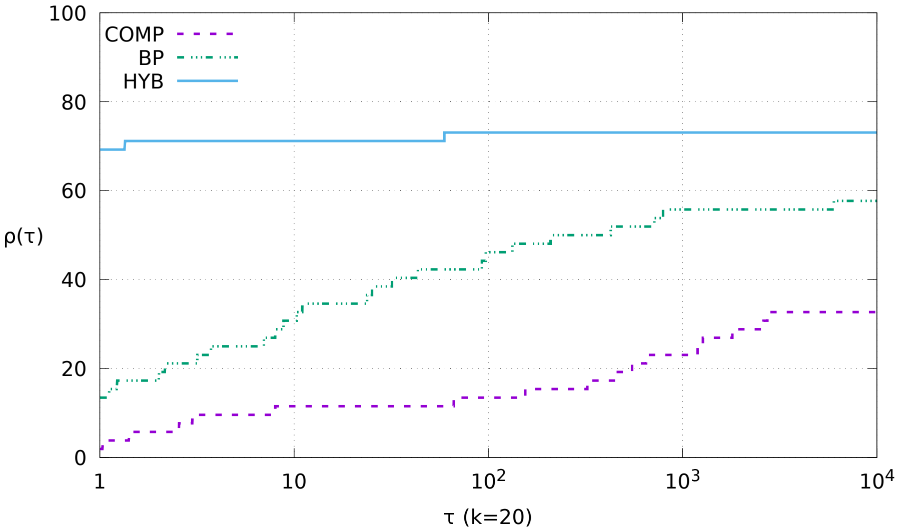
Computational Experiments

Case with $k = 15$



Computational Experiments

Case with $k = 20$



Conclusions and future work

- Our hybrid formulation outperforms both CPLEX and $B\&P$;
- It is a thin formulation, with $O(n)$ variables
- We partially exploit a **hereditary property** on G (if a subset of edges is cycle-free, any subset of it is cycle-free too) to convexify $\Phi(x)$
- This allows us to derive an ILP formulation in the natural space (was open for some time)
- Where else can we exploit similar ideas?

Thank you for your attention.